



Е. И. Стенина

# ОСНОВЫ НАУЧНЫХ ИССЛЕДОВАНИЙ

Определение статистических  
показателей выборки

Екатеринбург  
2020

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное  
учреждение высшего образования  
«Уральский государственный лесотехнический университет»  
(УГЛТУ)

Кафедра автоматизации и инновационных технологий

Е. И. Стенина

# ОСНОВЫ НАУЧНЫХ ИССЛЕДОВАНИЙ

Определение статистических  
показателей выборки

Методические указания по выполнению практических,  
лабораторных и исследовательских работ обучающимися  
по направлению 35.03.02 «Технология лесозаготовительных  
и деревоперерабатывающих производств»  
всех форм обучения

Екатеринбург  
2020

Печатается по рекомендации методической комиссии ИЛБ  
ФГБОУ ВО «Уральский государственный лесотехнический университет».  
Протокол № 2 от 03 октября 2019 г.

Рецензент – профессор кафедры АИТ, д-р техн. наук Гороховский А. Г.

Редактор Е. Л. Михайлова  
Оператор компьютерной верстки Т. В. Упорова

---

Подписано в печать 07.04.20		Поз. 18
Плоская печать	Формат 60×84/16	Тираж 10 экз.
Заказ №	Печ. л. 0,7	Цена руб. коп.

---

Редакционно-издательский отдел УГЛТУ  
Сектор оперативной полиграфии УГЛТУ

## ВВЕДЕНИЕ

Современное производство требует от специалиста умения самостоятельно ставить и решать различные принципиально новые задачи. Этого нельзя достичь без овладения основами научных исследований. В лесной и деревообрабатывающей промышленности исследования проводят зачастую с целью отыскания выгодных условий протекания процессов, оптимальных режимов работы и параметров машин и механизмов при их модернизации, а также состава многокомпонентных систем, оптимального размещения предприятий и их рациональной структуры в зависимости от района расположения и т. п.

Любое экспериментальное исследование условно можно разделить на три этапа: подготовка эксперимента, планирование и постановка опытов, а также обработка результатов измерений и их анализ.

Как правило, подготовка эксперимента заключается в уточнении целей и задач научно-исследовательской работы (НИР), составлении аналитического обзора, в выборе и обосновании направления исследований, а также в подготовке к проведению НИР.

Зачастую постановке основного эксперимента предшествует проведение разведывательных опытов, результаты которых подвергают статистической обработке с целью определения необходимого числа повторений каждого опыта, проводимого в рамках НИР, с целью обеспечения необходимой точности и достоверности ее результатов.

## 1. ЦЕЛЬ РАБОТЫ

Методические указания предназначены для приобретения обучающимися практических навыков в статистической обработке результатов разведывательных экспериментов.

## 2. СТАТИСТИЧЕСКАЯ ОБРАБОТКА РЕЗУЛЬТАТОВ ЭКСПЕРИМЕНТОВ

**Опыт** – часть эксперимента, выполненная при определенных значениях одного или нескольких факторов. С целью снижения вероятности ошибки при анализе результатов эксперимента необходимо дублирование каждого опыта.

**Выходным фактором** называется результат эксперимента (наблюдений)  $y_i$ , который является случайной величиной, так как всегда в большей или меньшей степени содержит ошибки, обусловленные погрешностью приборов, измерений, расчетов и т.п.

**Математическая статистика** – это наука о математических методах обработки, систематизации и использовании результатов наблюдений для научных и практических выводов.

Множество значений случайной величины, полученных в результате дублирования эксперимента или наблюдений над объектом исследования, представляет собой **статистическую совокупность**.

Статистическая совокупность, содержащая в себе все возможные значения случайной величины, называется **генеральной статистической совокупностью**.

**Выборочной статистической совокупностью (выборкой)** называется совокупность, в которой содержится только некоторая часть элементов генеральной совокупности.

**Основными статистическими показателями**, используемыми для характеристики любой выборочной совокупности (выборки), являются **объем выборки, выборочное среднее и выборочная дисперсия**.

Число опытов (наблюдений)  **$n$** , содержащееся в выборке, называется **объемом выборки**. Чем больше объем выборки, тем меньше влияние случайных ошибок, тем большей достоверностью обладают результаты, содержащиеся в выборке.

Наилучшей оценкой истинного значения выходной величины является **выборочное среднее  $\bar{y}$** .

Количественной оценкой модуля величины случайной ошибки измерения, возникающей под действием факторов, неизвестных исследователю или известных, но нерегулируемых и постоянно изменяющихся, является **выборочная дисперсия  $S^2$** .

### 3. ОБЩИЕ УКАЗАНИЯ ПО ВЫПОЛНЕНИЮ РАБОТЫ

При обработке результатов эксперимента (выборки):

- рассчитать статистические показатели выборки (п. п. 4.1 – 4.7);
- проанализировать результаты расчетов (п. 4.8).

Расчеты должны выполняться и **оформляться в соответствии с требованиями ЕСКД**.

### 4. ПОРЯДОК ВЫПОЛНЕНИЯ РАБОТЫ

С целью демонстрации методики статистической обработки результатов эксперимента будем использовать следующий пример.

#### *Пример*

По результатам дублированных опытов по антисептированию сосновых образцов были получены следующие значения удержания защитного средства  $y_i$ , г/м<sup>2</sup>: 115,30; 74,04; 119,22; 138,82; 65,30.

#### 4.1. Расчет среднего выборочного

Самым известным вошедшим в практику вариационно-статистическим элементом является среднее выборочное, или просто «среднее», которое рассчитывается по формуле (1). Найденное  $\bar{y}$  называют также оценкой математического ожидания.

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n}, \quad (1)$$

где  $y_i$  – наблюдения, содержащиеся в выборке;  
 $n$  – объем выборки.

##### *Пример*

$$\bar{y} = \frac{115,30 + 74,04 + 119,22 + 138,82 + 65,30}{5} = 102,54 \text{ г/м.}^2$$

#### 4.2. Расчет выборочных дисперсии и стандартного отклонения

Количественной оценкой величины случайных ошибок исследования является выборочная дисперсия  $S^2$ , а также выборочное стандартное отклонение (выборочный стандарт)  $S$ , которые выражаются в единицах того же наименования, что и среднее арифметическое.

Выборочная дисперсия рассчитывается по формуле (2) или (3), а выборочный стандарт – по формуле (4).

$$S^2 = \frac{\sum_{i=1}^n y_i^2 - n\bar{y}^2}{n-1}. \quad (2)$$

В случае, когда в выборке содержатся величины с большим абсолютным значением, лучше для расчета выборочной дисперсии использовать формулу (3).

$$S^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}, \quad (3)$$

$$S = \sqrt{S^2}. \quad (4)$$

**Пример**

$$S^2 = \frac{115,30^2 + 74,04^2 + 119,22^2 + 138,82^2 + 65,30^2 - 5 \times 102,54^2}{4} = 988,69 \text{ г/м}^2$$

или

$$S^2 = \frac{(115,30-102,54)^2 + (74,04-102,54)^2 + (119,22-102,54)^2 + (138,82-102,54)^2 + (65,30-102,54)^2}{5-1},$$

$$S = \sqrt{988,68} = 31,44 \text{ г/м}^2.$$

### 4.3. Расчет коэффициента вариации

При решении вопроса об изменчивости того или иного свойства необходимо вычислить коэффициент вариации  $\nu$ , который характеризует относительное рассеивание случайной величины от выборочного среднего:

$$\nu = \frac{S}{\bar{y}} 100\%. \quad (5)$$

**Пример**

$$\nu = \frac{31,44}{102,54} 100 = 30,67\%.$$

### 4.4. Расчет средней квадратической ошибки выборочного среднего

Величину среднего выборочного всегда находят из сравнительно небольшого количества наблюдений, так как измерить все отдельные значения интересующего свойства невозможно и не нужно. Поэтому необходимо иметь дополнительную характеристику, которая позволила бы по частному значению среднего выборочного судить об общей величине среднего арифметического изучаемой выходной величины. Такого рода характеристикой является средняя квадратическая ошибка среднего арифметического:

$$S_{\bar{y}} = \frac{S}{\sqrt{n}}. \quad (6)$$

**Пример**

$$S_{\bar{y}} = \frac{31,44}{\sqrt{5}} = 14,06 \text{ г/м}^2.$$

#### 4.5. Расчет показателя точности среднего значения

Подобно вариационному коэффициенту средняя квадратическая ошибка может быть выражена в процентах от соответствующего ей среднего арифметического. Полученная величина называется показателем точности среднего значения:

$$\xi = \frac{S_{\bar{y}}}{\bar{y}} 100. \quad (7)$$

Чем меньше показатель точности, тем надежнее результаты исследования. Принято считать, что в области лесной и деревообрабатывающей промышленности достаточная надежность будет обеспечена только в том случае, если показатель точности не превышает 5 %.

##### *Пример*

$$\xi = \frac{14,06}{102,54} 100 = 13,71\%.$$

Полученный показатель точности превышает 5 %, следовательно, можно сделать вывод о том, что результаты исследования недостаточно надежны. Для повышения их надежности необходимо увеличить объем выборки.

#### 4.6. Расчет доверительного интервала для математического ожидания

Выборочное среднее арифметическое  $\bar{y}$  представляет ценность постольку, поскольку по нему можно судить об истинном среднем, генеральном среднем, или математическом ожидании  $M_y$ . Представляет интерес отыскание величины максимальной ошибки  $\Delta$ , которую допускают, предполагая, что  $M_y = \bar{y}$ . Поэтому требуется найти величину  $\Delta$ , при которой выполняется условие

$$\bar{y} - \Delta \leq M_y \leq \bar{y} + \Delta, \quad (8)$$

$$\Delta = \frac{t_{табл} S}{\sqrt{n}}. \quad (9)$$

Однако нельзя абсолютно достоверно указать величину  $\Delta$ , так как расчет этой величины, как и любой статистический вывод, делают на основе результата эксперимента, а он заведомо содержит ошибки. Выводы, которые делают на основе неточных данных, принципиально не могут быть



абсолютно достоверными. Поэтому говорят о надежности статистического вывода, которую оценивают **уровнем значимости**  $q$  (в деревообработке принимается равным 0,05). Таким образом, из таблицы распределения Стьюдента (приложение) по уровню значимости  $q$  и числу степеней свободы  $f = n - 1$  находят  $t_{табл}$ .

### **Пример**

Для уровня значимости  $q = 0,05$  и числа степеней свободы  $f = 5 - 1 = 4$   $t_{табл} = 2,78$  (см. приложение).

$$\Delta = \frac{2,78 \times 31,44}{\sqrt{5}} = 39,09 \text{ г/м}^2,$$

$$102,54 - 39,09 \leq M_y \leq 102,54 + 39,09,$$

$$63,45 \text{ г/м}^2 \leq M_y \leq 141,63 \text{ г/м}^2.$$

Из вышеизложенного можно сделать вывод, что истинное среднее значение удержания защитного средства на сосновой древесине при однократном нанесении находится в пределах от 63,45 до 141,63 г/м<sup>2</sup>.

## **4.7. Определение необходимого объема выборки**

Требуется найти минимальное число повторений опытов  $n$ , при котором среднее арифметическое  $\bar{y}$ , определенное по этой выборке, отличалось бы от математического ожидания не более чем на заданную величину максимальной ошибки  $\Delta$ . Чем больше  $n$ , тем меньше максимальная ошибка  $\Delta$ , а значит, показатель точности среднего значения  $\zeta$ . Поэтому необходимо задаться желаемым значением максимальной ошибки  $\Delta$ . Тогда искоемое значение  $n$  определяется по формуле

$$n = \frac{t_{табл}^2 S^2}{\Delta^2}. \quad (10)$$

### **Пример**

Учитывая, что показатель точности среднего значения  $\zeta = 13,71 \%$  (п. 4.5) более чем в 2 раза превышает рекомендуемое значение (5 %), задаемся меньшим значением допустимой максимальной ошибки, чем полученное по расчетам в п. 4.6 (39,09 г/м<sup>2</sup>). Пусть  $\Delta = 20 \text{ г/м}^2$ .

$$n = \frac{2,78^2 \cdot 988,69}{20^2} = 19,10.$$

Таким образом, для снижения максимальной ошибки в 2 раза необходимо увеличить объем выборки (число дублирований эксперимента) до 20.

#### 4.8. Анализ результатов эксперимента

В выводах необходимо:

- перечислить основные статистические показатели, характеризующие исследуемую выборку;
- указать, обладают ли результаты, содержащиеся в выборке, достаточной надежностью или нет (в последнем случае необходимо дать рекомендации);
- указать, в каком интервале находится истинное значение выходной величины;
- указать необходимое число дублирований опыта с целью повышения надежности результатов экспериментов.

##### *Пример*

Исследуемую выборку характеризуют следующие основные статистические показатели:

- выборочное среднее  $\bar{y} = 102,54 \text{ г/м}^2$ ;
- выборочная дисперсия  $S^2 = 988,69 \text{ г/м}^2$ ;
- объем выборки  $n = 5$ .

Результаты, содержащиеся в исследуемой выборке, не обладают достаточной надежностью, так как показатель точности среднего значения  $\zeta = 13,71 \%$  превышает рекомендуемое значение (5 %). С целью повышения надежности опытных данных необходимо увеличить объем выборки (число дублирований эксперимента) до 20.

## РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

1. Пижурин, А. А. Основы научных исследований в деревообработке : учебник для студентов вузов, обучающихся по дневной и заоч. форме специальностей 250403 (260200) «Технология деревообработки» и 150405 (170400) «Машины и оборудование лесного комплекса» / А. А. Пижурин, А. А. Пижурин ; Моск. гос. ун-т леса. – М. : МГУЛ, 2005. – 305 с.

2. Пижурин, А. А. Научные исследования в деревообработке: Основы научных исследований : текст лекций : для студентов специальностей 2602.00 и 1704.00 / А. А. Пижурин ; [ред. Е. Г. Петрова] ; Моск. гос. ун-т леса. – М. : МГУЛ, 1999. – 103 с.

## Приложение

Значения  $t$ -критерия Стьюдента  
( $q$  – уровень значимости,  $f$  – число степеней свободы)

$f$	$q$	$q$
	0,05	0,01
1	12,71	63,66
2	4,30	9,92
3	3,18	5,84
4	2,78	4,60
5	2,57	4,03
6	2,45	3,71
7	2,36	3,50
8	2,31	3,36
9	2,26	3,25
10	2,23	3,17
15	2,13	2,95
20	2,09	2,85
30	2,04	2,75
40	2,02	2,70
50	2,01	2,68
60	2,00	2,66
80	1,99	2,64
100	1,98	2,63
120	1,98	2,62
200	1,97	2,60
500	1,96	2,59
$\infty$	1,96	2,58